Backward Analysis via Over-Approximate Abstraction and Under-Approximate Subtraction

Technical Report

Alexey Bakhirkin[1], Josh Berdine[2], and Nir Piterman[1]

[1] University of Leicester, Department of Computer Science
[2] Microsoft Research

Abstract. We propose a novel approach for computing weakest liberal safe preconditions of programs. The standard approaches, which call for either under-approximation of a greatest fixed point, or complementation of a least fixed point, are often difficult to apply successfully. Our approach relies on a different decomposition of the weakest precondition of loops. We exchange the greatest fixed point for the computation of a least fixed point above a recurrent set, instead of the bottom element. Convergence is achieved using over-approximation, while in order to maintain soundness we use an under-approximating logical subtraction operation. Unlike general complementation, subtraction more easily allows for increased precision in case its arguments are related. The approach is not restricted to a specific abstract domain and we use it to analyze programs using the abstract domains of intervals and of 3-valued structures.

1 Introduction

Forward static analyses usually compute program invariants which hold of executions starting from given initial conditions, e.g., over-approximations of reachable states. Conversely, backward static analyses for universal properties compute program invariants which ensure given assertions hold of all executions, e.g., under-approximations of safe states. Forward analysis of programs has been a notable success, while such backward analysis has seen much less research and is done less frequently (a notable example is [16]).

The standard formulation of forward analyses involves over-approximating a least fixed point of a recursive system of equations (transformers) that over-approximate the forward semantics of commands. Conversely, backward analyses for universal properties usually involve under-approximating a greatest fixed point of under-approximate equations.

The over-approximating abstractions used by forward analyses are far more common and well-developed than the under-approximations used by backward analyses. One approach to under-approximation is via over-approximate abstraction and under-approximate complementation $\overline{(\cdot)}$. For example, lower widening $p \mathbin{\underline{\triangledown}} q$ may be seen as $\overline{\overline{p} \mathbin{\triangledown} \overline{q}}$. However, computing the complement is, in many cases, infeasible or impractical (e.g., for 3-valued structures [21], separation logic [7], or polyhedra [11]).

Here, we suggest an alternative backward analysis approach that uses least fixed-point approximation, and an *under-approximate logical subtraction* operation in lieu of complementation. (Logical subtraction can also be understood as *and with complement* or *not implies*.) We show how to extend a computation of a recurrent set of a program with a least fixed-point approximation to obtain an under-approximation of the safe states from which *no* execution can lead to a failure (such as violating an assertion, dividing by zero, or dereferencing a dangling-pointer – i.e., an event that causes program execution to immediately abort and signal an error). Soundness is ensured by subtracting an over-approximation of the unsafe states.

Using subtraction instead of complementation has several advantages. First, it is easier to define in power set domains for which complementation can be hard or impractical. Second, as the approximations of safe and unsafe states are the results of analyzing the same code, they are strongly related and so subtraction may be more precise than a general under-approximate complementation.

Our approach is not restricted to a specific abstract domain and we use it to analyze numeric examples (using the domain of intervals) and examples coming from shape analysis (using the domain of 3-valued structures).

2 Preliminaries

Let $\mathcal{U}$ denote the set of program *memory* states and $\epsilon \notin \mathcal{U}$ a *failure* state. The concrete domain for our analysis is the power set $\mathcal{P}(\mathcal{U})$ ordered by $\subseteq$, with least element $\varnothing$, greatest element $\mathcal{U}$, join $\cup$, and meet $\cap$.

We introduce an abstract domain $\mathcal{D}$ (with $\sqsubseteq$, $\bot$, $\top$, $\sqcup$, and $\sqcap$) and a concretization function $\gamma : \mathcal{D} \to \mathcal{P}(\mathcal{U})$. For an element of an abstract domain, $d \in \mathcal{D}$, $\gamma(d)$ is the set of states represented by it. For example, for a program with two variables x and y, an element of the interval domain $d = \langle x : [1; 2], y : [3; 4]\rangle$ represents all states satisfying $(1 \leq x \leq 2) \wedge (3 \leq y \leq 4)$, i.e., $\gamma(d) = \{(x, y) \mid 1 \leq x \leq 2 \wedge 3 \leq y \leq 4\}$.

For a lattice $\mathcal{L}$, we define *complementation* as a function $\overline{(\cdot)} \colon \mathcal{L} \to \mathcal{L}$ such that for every $l \in \mathcal{L}$, $\gamma(\bar{l}) \cap \gamma(l) = \varnothing$ (i.e., they represent disjoint sets of states – but we do not require that $\gamma(\bar{l}) \cup \gamma(l) = \mathcal{U}$). For example, if $d \in \mathcal{D}$ over-approximates the unsafe states, then $\overline{d}$ under-approximates the safe states. For our concrete domain $\mathcal{P}(\mathcal{U})$ (and similarly, for every power set of atomic elements), we can use standard set-theoretic complement: $\overline{S} = \mathcal{U} \setminus S$.

We define *subtraction* as a function $(\cdot - \cdot) \colon \mathcal{L} \to \mathcal{L} \to \mathcal{L}$ such that for $l_1, l_2 \in \mathcal{L}$ we have $\gamma(l_1 - l_2) \subseteq \gamma(l_1)$ and $\gamma(l_1 - l_2) \cap \gamma(l_2) = \varnothing$. For example, given a domain $\mathcal{D}$, we can define subtraction for the power set domain $\mathcal{P}(\mathcal{D})$ as

$$D_1 - D_2 = \{d_1 \in D_1 \mid \forall\, d_2 \in D_2.\ \gamma(d_1) \cap \gamma(d_2) = \varnothing\} \tag{1}$$

This way, subtraction can be defined in e.g., the domain of 3-valued structures that does not readily support complementation. We claim that a useful subtraction is often easier to define than a useful complementation. We also note that for every $l_0 \in \mathcal{L}$, the function $\lambda l.(l_0 - l)$ is a complementation. However, for a given l, the accuracy of this complement depends on the actual choice of l_0.

2.1 Programming Language Syntax and Semantics

We consider a simple structured programming language. Given a set of *atomic statements* A ranged over by a, statements C of the language are constructed as follows:

$$\begin{array}{lll} C ::= a & & \text{atomic statement} \\ \mid C_1 \,;\, C_2 & & \text{sequential composition: executes } C_1 \text{ and then } C_2 \\ \mid C_1 + C_2 & & \text{branch: non-deterministically branches to either } C_1 \text{ or } C_2 \\ \mid C^* & & \text{loop: iterated sequential composition of } \geq 0 \text{ copies of } C \end{array}$$

We assume A contains: the empty statement `skip`, an assertion statement $\texttt{assert}\,\varphi$ (for a state formula φ), and an assumption statement $[\varphi]$. Informally, an assertion immediately aborts the execution and signals an error if φ is not satisfied, and we consider that there are no *valid* executions violating assumptions. Standard conditionals $\texttt{if}(\varphi)\ C_1\ \texttt{else}\ C_2$ can be expressed by $([\varphi]; C_1) + ([\neg\varphi]; C_2)$. Similarly, loops $\texttt{while}(\varphi)\ C$ can be expressed by $([\varphi]; C)^* \,;\, [\neg\varphi]$.

A state formula φ denotes a set of non-failure states $[\![\varphi]\!] \subseteq \mathcal{U}$ that satisfy φ. The semantics of a statement C is a relation $[\![C]\!] \subseteq \mathcal{U} \times (\mathcal{U} \cup \{\epsilon\})$. For $s, s' \in \mathcal{U}$, $[\![C]\!](s, s')$ means that executing C in state s *may* change the state to s'. Then, $[\![C]\!](s, \epsilon)$ means that s is unsafe: executing C from state s may result in failure (may cause the program to immediately abort). Let $\Delta_{\mathcal{U}}$ be the diagonal relation on states $\Delta_{\mathcal{U}} = \{(s, s) \mid s \in \mathcal{U}\}$. Let composition of relations in $\mathcal{U} \times (\mathcal{U} \cup \{\epsilon\})$ be defined as $S \fatsemi R = (R \cup \{(\epsilon, \epsilon)\}) \circ S$ where $\circ$ is standard composition of relations. Fixed points in $\mathcal{U} \times (\mathcal{U} \cup \{\epsilon\})$ are with respect to the subset order, where $\mathrm{lfp}\, \lambda X.\, F(X)$ denotes the least fixed point of $\lambda X.\, F(X)$, and similarly, $\mathrm{gfp}\, \lambda X.\, F(X)$ denotes the greatest fixed point of $\lambda X.\, F(X)$. For an atomic statement a, we assume that $[\![a]\!]$ is a predefined left-total relation, and the semantics of other statements is defined as follows:

$$\begin{array}{rlrl} [\![\texttt{skip}]\!] &= \Delta_{\mathcal{U}} & [\![C_1 \,;\, C_2]\!] &= [\![C_1]\!] \fatsemi [\![C_2]\!] \\ [\![[\varphi]]\!] &= \{(s, s) \mid s \in [\![\varphi]\!]\} & [\![C_1 + C_2]\!] &= [\![C_1]\!] \cup [\![C_2]\!] \\ [\![\texttt{assert}\,\varphi]\!] &= \{(s, s) \mid s \in [\![\varphi]\!]\} \cup & [\![C^*]\!] &= \mathrm{lfp}\, \lambda X.\, \Delta_{\mathcal{U}} \cup ([\![C]\!] \fatsemi X) \\ & \quad \{(s, \epsilon) \mid s \in \mathcal{U} \wedge s \notin [\![\varphi]\!]\} & & \end{array}$$

Note that the assumption that atomic statements denote left-total relations excludes statements that affect control flow such as `break` or `continue`. In what follows, we constrain considered programs in the following way. Programs cannot have nested loops and assumption statements $[\varphi]$ are only allowed to appear at the start of branches and at the entry and exit of loops (they cannot be used as normal atomic statements):

$$C ::= a \mid C_1 \,;\, C_2 \mid ([\varphi] \,;\, C_1) + ([\psi] \,;\, C_2) \mid ([\psi] \,;\, C)^* \,;\, [\varphi]$$

We require that for branches and loops, $\varphi \vee \psi = 1$ (i.e., $[\![\varphi]\!] \cup [\![\psi]\!] = \mathcal{U}$). That is, for a loop-free statement, the domain of its semantics is $\mathcal{U}$. We also require that the language of state formulas is closed under negation.

2.2 Fixed-Point Characterizations of Safe and Unsafe States

Given a statement C and a set of states $S \subseteq \mathcal{U}$, we define:

- $pre(C, S) = \{s \in \mathcal{U} \mid \exists s' \in S.\ [\![C]\!](s, s')\}$. The states that may lead to S after executing C.
- $fail(C) = \{s \in \mathcal{U} \mid [\![C]\!](s, \epsilon)\}$. The *unsafe* states: those that may cause C to fail.
- $wp(C, S) = \{s \in \mathcal{U} \mid \forall s' \in \mathcal{U} \cup \{\epsilon\}.\ [\![C]\!](s, s') \Rightarrow s' \in S\}$. The *weakest liberal precondition* that ensures safety [6] – safe states that must lead to S if execution of the statement terminates.

We abbreviate $pre(C, S) \cup fail(C)$ to $pre{+}fail(C, S)$.

Lemma 1. *For a statement C and a set of states $S \subseteq \mathcal{U}$,* $wp(C, S) = \mathcal{U} \smallsetminus pre{+}fail(C, \mathcal{U} \smallsetminus S)$.

The proof is a direct calculation based on the definitions. See Appendix A for proofs.

For a program C, our goal is to compute (an under-approximation of) $wp(C, \mathcal{U})$, and (an over-approximation of) its complement $fail(C)$. If we are interested in termination with specific postcondition φ, we add an $\texttt{assert}\,\varphi$ statement to the end of the program. We characterize these sets (as is standard [8,9]) as solutions to two functionals P and N that associate a statement C and a set of states S (resp., V) $\subseteq \mathcal{U}$ with a predicate $P(C, S)$, resp., $N(C, V)$. $P(C, S)$ (the *positive side*) denotes the states that must either lead to successful termination in S or cause non-termination, and $N(C, V)$ (the *negative side*) denotes the states that may lead to failure or termination in V.

$$
\begin{array}{ll}
P(a, S) = wp(a, S) & N(a, V) = pre{+}fail(a, V) \\
P([\varphi], S) = [\![\neg\varphi]\!] \cup S & N([\varphi], V) = [\![\varphi]\!] \cap V \\
P(\texttt{assert}\,\varphi, S) = [\![\varphi]\!] \cap S & N(\texttt{assert}\,\varphi, V) = [\![\neg\varphi]\!] \cup V \\
P(C_1; C_2, S) = P(C_1, P(C_2, S)) & N(C_1; C_2, V) = N(C_1, N(C_2, V)) \\
P(C_1 + C_2, S) = P(C_1, S) \cap P(C_2, S) & N(C_1 + C_2, V) = N(C_1, V) \cup N(C_2, V) \\
P(C^*, S) = \mathrm{gfp}\,\lambda X.\, S \cap P(C, X) & N(C^*, V) = \mathrm{lfp}\,\lambda Y.\, V \cup N(C, Y)
\end{array}
$$

Lemma 2. *For a statement C and set of states $S \subseteq \mathcal{U}$, $P(C, S) = \mathcal{U} \smallsetminus N(C, \mathcal{U} \smallsetminus S)$.*

The proof is by structural induction.

Lemma 3. *For a statement C and sets of states $S, V \subseteq \mathcal{U}$, $P(C, S) = wp(C, S)$, and $N(C, V) = pre{+}fail(C, V)$.*

The proof is by structural induction, relying on continuity of *pre+fail*.

3 Least Fixed-Point Characterization of Safe States

The direct solution of the positive side is by under-approximating a greatest fixed point. This can be problematic since most domains are geared towards over-approximating least fixed points. Hence, we are not going to approximate the greatest fixed point for the positive side directly. Instead, we restate the problem for loops such that the resulting characterization leads to a least fixed point computation where termination is ensured by using an appropriate over-approximate abstraction.

In this section, we focus on the looping statement:

$$C_{\text{loop}} = ([\psi]\ ;\ C_{\text{body}})^*\ ;\ [\varphi] \tag{2}$$

where C_{body} is the *loop body*; if ψ holds the execution may enter the loop body; and if φ holds the execution may exit the loop. To simplify the presentation, in what follows, we assume that the semantics of C_{body} is directly known. Since C_{body} is itself loop-free, $[\![C_{\text{body}}]\!]$ does not induce fixed points, and the transformers for the loop body can be obtained, e.g., by combining the transformers for its sub-statements.

3.1 Recurrent Sets

We reformulate the characterizations of safe states in terms of least fixed points with the use of recurrent sets. For the loop in (2), an *existential recurrent set* is a set $R_\exists$, s.t.

$$\begin{aligned} &R_\exists \subseteq [\![\psi]\!] \\ &\forall s \in R_\exists.\ \exists s' \in R_\exists.\ [\![C_{\text{body}}]\!](s, s') \end{aligned}$$

These are states that may cause non-termination (i.e., cause the computation to stay inside the loop forever). For the loop in (2), a *universal recurrent set* is a set $R_\forall$, s.t.

$$\begin{aligned} &R_\forall \subseteq [\![\neg\varphi]\!] \\ &\forall s \in R_\forall.\ \big(\forall s' \in \mathcal{U} \cup \{\epsilon\}.\ [\![C_{\text{body}}]\!](s, s') \Rightarrow s' \in R_\forall\big) \end{aligned}$$

These are states that must cause non-termination. For practical reasons discussed later in Sect. 4.2, we *do not* require these sets to be maximal.

Lemma 4. *For the loop* $C_{\text{loop}} = ([\psi]\ ;\ C_{\text{body}})^*\ ;\ [\varphi]$*, and a set of states* $S \subseteq \mathcal{U}$

$$R_\forall \subseteq P(C_{\text{loop}}, S) \qquad\qquad R_\exists \smallsetminus N(C_{\text{loop}}, \mathcal{U} \smallsetminus S) \subseteq P(C_{\text{loop}}, S)$$

For $R_\forall$, the proof correlates universal recurrence and *wp*, relying on monotonicity of $P(C_{\text{loop}}, \cdot)$. For $R_\exists$, the result follows from Lemma 2.

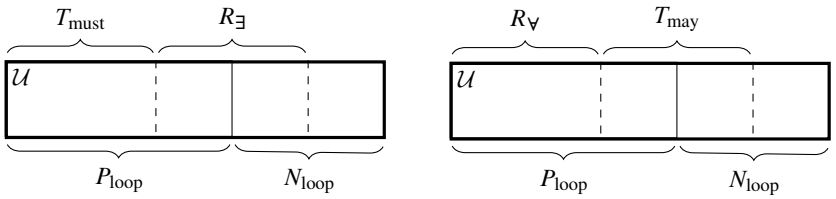
3.2 Positive Least Fixed Point via Recurrent Sets

We begin with an informal explanation of how we move from a greatest fixed point formulation to a least fixed point one. Observe that for the loop in (2), the positive and negative sides (following the definition in Sect. 2.2) are characterized by:

$$\begin{aligned} P(C_{\text{loop}}, S) &= \operatorname{gfp} \lambda X.\ ([\![\neg\varphi]\!] \cup S) \cap \big([\![\neg\psi]\!] \cup P(C_{\text{body}}, X)\big) \\ N(C_{\text{loop}}, V) &= \operatorname{lfp} \lambda Y.\ ([\![\varphi]\!] \cap V) \cup \big([\![\psi]\!] \cap N(C_{\text{body}}, Y)\big) \end{aligned} \tag{3}$$

Then, since loops only occur at the top level, a program C_{prg} that contains the loop C_{loop} can be expressed as $C_{\text{init}}\ ;\ C_{\text{loop}}\ ;\ C_{\text{rest}}$ (where C_{init} or C_{rest} may be `skip`). Let:

- $P_{\text{rest}} = P(C_{\text{rest}}, \mathcal{U})$ – the safe states of the loop's continuation.
- $N_{\text{rest}} = N(C_{\text{rest}}, \varnothing)$ – states that may cause failure of the loop's continuation. Note that $N_{\text{rest}} = \mathcal{U} \smallsetminus P_{\text{rest}}$.



(a) Partitioning with existential recurrence. (b) Partitioning with universal recurrence.

Fig. 1: Partitioning of the states at the loop entry.

- $P_{\text{loop}} = P(C_{\text{loop}}, P_{\text{rest}})$ – the safe states of the loop and its continuation.
- $N_{\text{loop}} = N(C_{\text{loop}}, N_{\text{rest}})$ – states that may cause failure of the loop or its continuation. Note that $N_{\text{loop}} = \mathcal{U} \smallsetminus P_{\text{loop}}$.

For the loop in (2), Fig. 1 shows how the states entering the loop can be partitioned. In the figure, by T_{must}, we denote the states that must cause successful termination *of the loop* (in a state belonging to P_{rest}), and by T_{may}, we denote states that may cause successful termination.

Fig. 1a shows that the positive side for the loop in (2) can be partitioned into the following two parts:

- $R_\exists \smallsetminus N_{\text{loop}}$ – states that may cause non-termination but may not fail;
- T_{must} – states that must cause successful termination of the loop.

T_{must} can be characterized as the least fixed point:

$$T_{\text{must}} = \operatorname{lfp} \lambda X.\, (\llbracket \neg\psi \rrbracket \cap P_{\text{rest}}) \cup \Big(((\llbracket \psi \rrbracket \cap P_{\text{rest}}) \cup \llbracket \neg\varphi \rrbracket) \cap wp(C_{\text{body}}, X) \Big)$$

Intuitively, the states in $\llbracket \neg\psi \rrbracket \cap P_{\text{rest}}$ cause the loop to immediately terminate (such that the rest of the program does not fail), those in $((\llbracket \psi \rrbracket \cap P_{\text{rest}}) \cup \llbracket \neg\varphi \rrbracket) \cap wp(C_{\text{loop}}, \llbracket \neg\psi \rrbracket \cap P_{\text{rest}})$ can make one iteration through the loop, and so on.

Fig. 1b shows that the positive side can also be partitioned in another way:

- $R_\forall$ – states that must cause non-termination of the loop;
- $T_{\text{may}} \smallsetminus N_{\text{loop}}$ – states that may cause successful termination but may not fail.

In a way similar to [9], T_{may} can be characterized as the least fixed point:

$$T_{\text{may}} = \operatorname{lfp} \lambda X.\, (\llbracket \varphi \rrbracket \cap P_{\text{rest}}) \cup \big(\llbracket \psi \rrbracket \cap pre(C_{\text{body}}, X) \big)$$

Intuitively, from states $\llbracket \varphi \rrbracket \cap P_{\text{rest}}$, the loop *may* immediately terminate in a state safe for C_{rest}, from states $\llbracket \psi \rrbracket \cap pre(C_{\text{body}}, \llbracket \varphi \rrbracket \cap P_{\text{rest}})$ the loop may make one iteration and terminate, and so on. From this, it can be shown that

$$T_{\text{may}} \smallsetminus N_{\text{loop}} = \operatorname{lfp} \lambda X.\, \big((\llbracket \varphi \rrbracket \cap P_{\text{rest}}) \smallsetminus N_{\text{loop}} \big) \cup \big((\llbracket \neg\varphi \rrbracket \cap pre(C_{\text{body}}, X)) \smallsetminus N_{\text{loop}} \big)$$

We replace ψ with $\neg\varphi$, since the states in $\llbracket \psi \rrbracket \cap \llbracket \varphi \rrbracket \cap pre(C_{\text{body}}, X)$ are either already included in the first disjunct (if belonging to P_{rest}), or are unsafe and removed by subtraction.

Following these least fixed point characterizations, we re-express the equation for the positive side of the loop (3) using the existential recurrent set $R_\exists$ as follows, where $N = N(C_{\text{loop}}, \mathcal{U} \smallsetminus S)$:

$$P^\exists(C_{\text{loop}}, S) = \text{lfp}\, \lambda X.\, (R_\exists \smallsetminus N) \cup ([\![\neg\psi]\!] \cap S) \cup \Big(\big(([\![\psi]\!] \cap S) \cup [\![\neg\varphi]\!]\big) \cap wp(C_{\text{body}}, X)\Big) \tag{4}$$

or using the universal recurrent set $R_\forall$ as follows:

$$P^\forall(C_{\text{loop}}, S) = \text{lfp}\, \lambda X.\, R_\forall \cup \big([\![\varphi]\!] \cap S) \smallsetminus N\big) \cup \Big(\big([\![\neg\varphi]\!] \cap pre(C_{\text{body}}, X)\big) \smallsetminus N\Big) \tag{5}$$

Theorem 1. *The alternative characterizations of the positive side of the loop:* (4) *and* (5) *– under-approximate the original characterization* (3)*. That is, for a set* $S \subseteq \mathcal{U}$*,*

$$P^\exists(C_{\text{loop}}, S) \subseteq P(C_{\text{loop}}, S) \qquad P^\forall(C_{\text{loop}}, S) \subseteq P(C_{\text{loop}}, S)$$

4 Approximate Characterizations

In Sects. 2.2 and 3.2, we characterized both the negative and the positive sides as least fixed points. For the negative side, our goal is to over-approximate the least fixed point, and we can do that using standard tools. That is, we move to an abstract domain $\mathcal{D}(\sqsubseteq, \bot, \top, \sqcup, \sqcap, \triangledown)$ where widening $\triangledown$ and join $\sqcup$ may coincide for domains that do not allow infinite ascending chains. For the positive side our goal is to under-approximate the least fixed point, and to do so, we build an increasing chain of its approximations and use the previously computed negative side and subtraction to ensure soundness.

As before, since we do not allow nested loops, we assume that abstract transformers for loop bodies are given. For a loop-free statement C and $d \in \mathcal{D}$, we assume: over- and under-approximating transformers $pre^\sharp(C, d)$ and $wp^\flat(C, d)$, over-approximating operation $fail^\sharp(C)$; and for assumption statements $[\varphi]$: under- and over-approximate transformers $[\varphi, d]^\flat$ and $[\varphi, d]^\sharp$ such that:

$$\gamma(pre^\sharp(C, d)) \supseteq pre(C, \gamma(d)) \qquad \gamma(fail^\sharp(C)) \supseteq fail(C)$$
$$\gamma(wp^\flat(C, d)) \subseteq wp(C, \gamma(d)) \qquad \gamma([\varphi, d]^\flat) \subseteq [\![\varphi]\!] \cap \gamma(d) \subseteq \gamma([\varphi, d]^\sharp)$$

We abbreviate $[\varphi, \top]^\flat$ to $[\varphi]^\flat$ and $[\varphi, \top]^\sharp$ to $[\varphi]^\sharp$.

Note that the above includes both over-approximating and under-approximating operations. In section 4.1, we relax the requirements and obtain an analysis where subtraction is the only under-approximating operation.

For a statement C and $n \in \mathcal{D}$, the approximate negative side $N^\sharp(C, n)$, which over-approximates $N(C, \gamma(n))$, is (non-recursively) defined as follows:

$$N^\sharp(a, n) = pre \dot{+} fail^\sharp(a, n)$$
$$N^\sharp(C_1\,;\, C_2, n) = N^\sharp\big(C_1, N^\sharp(C_2, n)\big)$$

$$N^\sharp\big(([\varphi];C_1)+([\psi];C_2),n\big) = [\varphi,N^\sharp(C_1,n)]^\sharp \sqcup [\psi,N^\sharp(C_2,n)]^\sharp$$
$$N^\sharp\big(([\psi];C_{\text{body}})^*;[\varphi],n\big) = \text{the first } n_j \in \{n_i\}_{i\geq 0} \text{ such that } n_{j+1}\sqsubseteq n_j \text{ where}$$
$$n_0 = [\varphi,n]^\sharp \text{ and } n_{i+1} = n_i \triangledown [\psi,N^\sharp(C_{\text{body}},n_i)]^\sharp$$

For a statement C and a pair of elements $p,n\in\mathcal{D}$ that are disjoint ($\gamma(p)\cap\gamma(n)=\varnothing$), we define the approximate positive side $P^\flat(C,p,n)$ such that it under-approximates $P(C,\mathcal{U}\smallsetminus\gamma(n))$. $P^\flat(C,p,n)$ is defined mutually with an auxiliary $Q^\natural(C,p,n)$ by induction on the structure of C. Optimally, $Q^\natural(C,p,n)$ represents a tight under-approximation of $P(C,\gamma(p))$, but actually need not be an under-approximation. Also, note how n is used to abstractly represent the complement of the set of interest.

For loop-free code, $P^\flat$ and $Q^\natural$ are (non-recursively) defined as follows:

$$P^\flat(C,p,n) = Q^\natural(C,p,n) - N^\sharp(C,n)$$
$$Q^\natural(a,p,n) = wp^\flat(a,p)$$
$$Q^\natural(C_1\,;C_2,p,n) = P^\flat\big(C_1,P^\flat(C_2,p,n),N^\sharp(C_2,n)\big)$$
$$Q^\natural\big(([\varphi];C_1)+([\psi];C_2),p,n\big) = \big(P^\flat(C_1,p,n)\sqcap P^\flat(C_2,p,n)\big)\sqcup [\neg\psi,P^\flat(C_1,p,n)]^\flat\sqcup[\neg\varphi,P^\flat(C_2,p,n)]^\flat$$

For a loop $C_{\text{loop}}=([\psi];C_{\text{body}})^*;[\varphi]$, we define a sequence $\{q_i\}_{i\geq 0}$ of approximants to $Q^\natural(C_{\text{loop}},p,n)$, where $q_{i+1}=q_i\triangledown\tau(q_i)$ and the initial point q_0 and the transformer τ are defined following either the characterization (4) using an approximation $R^\natural_\exists\in\mathcal{D}$ of an existential recurrent set of the loop:

$$q_0 = \big(R^\natural_\exists - N^\sharp(C_{\text{loop}},n)\big)\sqcup[\neg\psi,p]^\flat$$
$$\tau(q_i) = \big([\psi,p]^\flat\sqcap wp^\flat(C_{\text{body}},q_i)\big)\sqcup[\neg\varphi,wp^\flat(C_{\text{body}},q_i)]^\flat$$

or the characterization (5) using an approximation $R^\natural_\forall\in\mathcal{D}$ of a universal recurrent set:

$$q_0 = R^\natural_\forall\sqcup\big([\varphi,p]^\flat - N^\sharp(C_{\text{loop}},n)\big)$$
$$\tau(q_i) = \big([\neg\varphi,pre^\sharp(C_{\text{body}},q_i)]^\flat - N^\sharp(C_{\text{loop}},n)\big)$$

As for loop-free commands, $Q^\natural$ can be computed first, and $P^\flat$ defined using the result. That is, define $Q^\natural(C_{\text{loop}},p,n)=q_j$ where p_j is the first element such that $q_{j+1}\sqsubseteq q_j$, and then define $P^\flat(C_{\text{loop}},p,n)=Q^\natural(C_{\text{loop}},p,n)-N^\sharp(C_{\text{loop}},n)$.

Alternatively, $P^\flat$ and $Q^\natural$ can be computed simultaneously by also defining a sequence $\{p_i\}_{i\geq 0}$ of safe under-approximants of $P^\flat(C_{\text{loop}},p,n)$, where $p_0 = q_0 - N^\sharp(C_{\text{loop}},n)$ and $p_{i+1}=(p_i\triangledown\tau(q_i))-N^\sharp(C_{\text{loop}},n)$. Then $P^\flat(C_{\text{loop}},p,n)=p_j$ where p_j is the first element such that $q_{j+1}\sqsubseteq q_j$ or $p_{j+1}\not\sqsupseteq p_j$. In this case, we may obtain a sound $P^\flat$ before the auxiliary $Q^\natural$ has stabilized. While we have not yet done rigorous experimental validation, we prefer this approach when dealing with coarse subtraction.

When analyzing a top-level program C_{prg}, the analysis starts with $N^\sharp(C_{\text{prg}},\bot)$ and precomputes $N^\sharp$ (an over-approximation of unsafe states) for all statements of the program. Then it proceeds to compute $P^\flat(C_{\text{prg}},\top,\bot)$ (an under-approximation of safe input states) reusing the precomputed results for $N^\sharp$.

Note that we are using join and widening on the positive side which means that $Q^\natural$ may not under-approximate the positive side of the concrete characterization. The use of widening allows for the ascending chain to converge, and subtraction of the negative side ensures soundness of $P^\flat$. In other words, while the alternate concrete characterizations (4) and (5) are used to guide the definition of the approximate characterizations, soundness is argued directly rather than by using (4) and (5) as an intermediate step.

Theorem 2. *For a statement C and $p, n \in \mathcal{D}$ s.t. $\gamma(p) \cap \gamma(n) = \varnothing$, $N^\sharp(C, n) \supseteq N(C, \gamma(n))$ and $P^\flat(C, p, n) \subseteq P(C, \mathcal{U} \smallsetminus \gamma(n))$. Hence, for a top-level program C_{prg}, $\gamma(N^\sharp(C_{\text{prg}}, \bot)) \supseteq N(C_{\text{prg}}, \varnothing)$ (i.e., it over-approximates input states that may lead to failure), and $\gamma(P^\flat(C_{\text{prg}}, \top, \bot)) \subseteq P(C_{\text{prg}}, \mathcal{U})$ (i.e., it under-approximates safe input states).*

The argument for $N^\sharp$ proceeds in a standard way [10]. Soundness for $P^\flat$ then follows due to the use of subtraction.

4.1 Optimizations of Constraints

Use of over-approximate operations Since we are subtracting $N^\sharp(C, n)$ anyway, we can relax the right-hand side of the definition of $Q^\natural(C, p, n)$ without losing soundness. Specifically, we can replace under-approximating and must- operations by their over-approximating and may- counterparts. This way, we obtain an analysis where subtraction is the only under-approximating operation.

- For a loop-free statement C, always use $pre^\sharp(C, p)$ in place of $wp^\flat(C, p)$ (note that we *already* use $pre^\sharp$ on the positive side for loop bodies when starting from a universal recurrent set). This can be handy, e.g., for power set domains where $pre^\sharp$ (unlike $wp^\flat$) can be applied element-wise. Also, these transformers may coincide for deterministic loop-free statements (if the abstraction is precise enough). Later, when discussing Example 2, we note some implications of this substitution.
- For a state formula φ, use $[\varphi, \cdot]^\sharp$ in place of $[\varphi, \cdot]^\flat$. Actually, for some combinations of an abstract domain and a language of formulas, these transformers coincide. For example, in a polyhedral domain, conjunctions of linear constraints with non-strict inequalities have precise representations as domain elements.
- For branching statements, use $[\varphi, P^\flat(C_1, p, n)]^\sharp \sqcup [\psi, P^\flat(C_2, p, n)]^\sharp$ in place of the original expression.
- In the definition of $Q^\natural$, an over-approximate meet operation $\sqcap^\sharp$ suffices.

The result of these relaxations is:

$$\begin{aligned} Q^\natural(a, p, n) &= pre^\sharp(a, p) \\ Q^\natural(C_1 \,;\, C_2, p, n) &= P^\flat\big(C_1, P^\flat(C_2, p, n), N^\sharp(C_2, n)\big) \\ Q^\natural\big(([\varphi]; C_1) + ([\psi]; C_2), p, n\big) &= [\varphi, P^\flat(C_1, p, n)]^\sharp \sqcup [\psi, P^\flat(C_2, p, n)]^\sharp \end{aligned}$$

$$\begin{aligned} q_0 &= \big(R^\natural_\exists - N^\sharp(C_{\text{loop}}, n)\big) \sqcup [\neg\psi, p]^\sharp \\ \tau(q_i) &= \big([\psi, p]^\sharp \sqcap^\sharp pre^\sharp(C_{\text{body}}, q_i)\big) \sqcup [\neg\varphi, pre^\sharp(C_{\text{body}}, q_i)]^\sharp \end{aligned}$$

or

$$q_0 = R^{\natural}_{\forall} \sqcup \left([\varphi, p]^\sharp - N^\sharp(C_{\text{loop}}, n)\right)$$
$$\tau(q_i) = \left([\neg\varphi, pre^\sharp(C_{\text{body}}, q_i)]^\sharp - N^\sharp(C_{\text{loop}}, n)\right)$$

No subtraction for $\mathbf{Q}^\natural$ For a similar reason, subtraction can be removed from the characterization of $Q^\natural$ without affecting soundness of $P^\flat$.

Bound on the positive side Another observation is that for a loop C_{loop} as in (2), the positive side $P(C_{\text{loop}}, S)$ is bounded by $[\![\neg\varphi]\!] \cup S$, as can be seen from the characterization (3). This can be incorporated into a specialized definition for loops, defining $P^\flat(C_{\text{loop}}, p, n) = \left(Q^\natural(C_{\text{loop}}, p, n) \sqcap ([\neg\varphi]^\sharp \sqcup p)\right) - N^\sharp(C_{\text{loop}}, n)$ or by performing the meet during computation of $Q^\natural$ by defining $q_{i+1} = \left(q_i \triangledown \tau(q_i)\right) \sqcap ([\neg\varphi]^\sharp \sqcup p)$.

4.2 Approximating the Recurrent Set

When approximating the positive side for a loop, the computation is initialized with an approximation of the recurrent set induced by the loop. Our analysis is able to start with either an existential or a universal recurrent set depending on what search procedure is available for the domain. The instantiation of our approach for numerical domains uses the tool E-HSF [4] that is capable of approximating both existential and universal recurrence. Other tools for numeric domains are described in [12,23]. The instantiation of our approach for the shape analysis with 3-valued logic uses a prototype procedure that we have developed to approximate existential recurrent sets.

Normally, the search procedures are incomplete: the returned sets only imply recurrence, and the search itself might not terminate (we assume the use of timeouts in this case). For this reason, in Sect. 3, we prefer not to define the recurrent sets to be maximal. This incompleteness leaves room for our analysis to improve the approximation of recurrence. For example, sometimes the solver produces a universal recurrence that is closed under forward propagation, but is not closed under backward propagation. In such cases, our analysis can produce a larger recurrent set.

5 Examples

In this section, we demonstrate our approach on several examples: first for a numeric domain, and then for the shape analysis domain of 3-valued structures. We note that numeric programs are considered here solely for the purpose of clarity of explanation, since the domain is likely to be familiar to most readers. We do not claim novel results specifically for the analysis of numeric programs, although we note that our approach may be able to complement existing tools. Detailed explanations of Examples are included in Appendix B.

Example 1 aims at describing steps of the analysis in detail (to the extent allowed by space constraints). Example 2 is restricted to highlights of the analysis and includes a pragmatic discussion on using $pre^\sharp$ on the positive side. Examples 3 and 4 consider programs from a shape analysis domain and we only report on the result of the analysis.

```
1  | while x ≥ 1 do
2  |   if x = 60 then
3  |     x ← 50
4  |   end
5  |   x ← x + 1
6  |   if x = 100 then
7  |     x ← 0
8  |   end
9  | end
10 | assert 0
```

(a) With syntactic sugar.

$$_1\big([x \geq 1];$$
$$\quad {}_2(([x = 60]; {}_3x \leftarrow 50) + ([x \neq 60]; \mathtt{skip}));$$
$$\quad {}_5x \leftarrow x + 1;$$
$$\quad {}_6(([x = 100]; {}_7x \leftarrow 0) + ([x \neq 100]; \mathtt{skip}));$$
$$\big)^*; [x \leq 0];$$
$$_{10}\,\mathtt{assert}\,0$$

(b) Desugared.

Fig. 2: Example program 1.

Example 1. In this example, we consider the program in Fig. 2: Fig. 2a shows program text using syntactic sugar for familiar while-language, and Fig. 2b shows the corresponding desugared program. We label the statements that are important for the analysis with the corresponding line numbers from Fig. 2a (like in ${}_3x \leftarrow 50$).

We assume that program variables (just x in this case) take *integer* values. For the abstract domain, we use disjunctive refinement over intervals allowing a bounded number of disjuncts (e.g., [2]). Recall that $\langle x : [a; b], y : [c; d]\rangle$ denotes a singleton abstract state of a program with two variables x and y, representing the set of concrete states, satisfying $(a \leq x \leq b) \wedge (c \leq y \leq d)$. Note that for this abstract domain and the formulas, appearing in the program, $[\cdot]^\flat$ and $[\cdot]^\sharp$ coincide, and we write $[\cdot]^\natural$ to denote either. To emphasize that the analysis can produce useful results even when using a coarse subtraction function, we use subtraction as defined in (1). That is, we just drop from the positive side those disjuncts that have a non-empty intersection with the negative side. For example, $\{\langle x : [1; 3]\rangle, \langle x : [5; 7]\rangle\} - \langle x : [6; 8]\rangle = \langle x : [1; 3]\rangle$. The analysis is performed mechanically by a prototype tool that we have implemented.

To simplify the presentation, in this example, we bound the number of disjuncts in a domain element by 2. Also to simplify the presentation, we omit the $\sharp$- and $\flat$-superscripts, and write, e.g., *pre+fail* for *pre+fail*$^\sharp$. For a statement labeled with i, we write N_i^j to denote the result of the j-th step of the computation of its negative side, and N_i to denote the computed value (similarly, for P).

We start with the analysis of the negative side. For the final statement,

$$N_{10}^1 = \textit{pre+fail}(\mathtt{assert}\,0, \bot) = \top$$

then, we proceed to the first approximation for the loop (for clarity, we compute *pre* of the body in steps),

$$N_1^1 = [x \leq 0, N_{10}^1]^\natural = \langle x : (-\infty; 0]\rangle$$
$$N_7^1 = \textit{pre+fail}(x \leftarrow 0, N_1^1) = \top$$
$$N_6^1 = [x = 100, N_7^1]^\natural \sqcup [x \neq 100, N_1^1]^\natural = \{\langle x : (-\infty; 0]\rangle, \langle x : [100]\rangle\}$$
$$N_5^1 = \textit{pre+fail}(x \leftarrow x + 1, N_6^1) = \{\langle x : (-\infty; -1]\rangle, \langle x : [99]\rangle\}$$

$$N_3^1 = \mathit{pre}\text{+}\mathit{fail}(x \leftarrow 50, N_5^1) = \bot$$
$$N_2^1 = [x = 60, N_3^1]^\sharp \sqcup [x \neq 60, N_5^1]^\sharp = \{\langle x : (-\infty; -1]\rangle, \langle x : [99]\rangle\}$$
$$N_1^2 = N_1^1 \sqcup [x \geq 1, N_2^1]^\sharp = \{\langle x : (-\infty; 0]\rangle, \langle x : [99]\rangle\}$$

then, repeating the same similar sequence of steps for the second time gives

$$N_1^2 = N_1^2 \sqcup [x \geq 1, N_2^2]^\sharp = \{\langle x : (-\infty; 0]\rangle, \langle x : [98, 99]\rangle\}$$

at which point we detect an unstable bound. The choice of widening strategy is not our focus here, and for demonstration purposes, we proceed without widening, which allows to discover the stable bound of 61. In a real-world tool, to retain precision, some form of widening *up to* [13] or landmarks [22] could be used. Thus, we take

$$N_1 = \{\langle x : (-\infty; 0]\rangle, \langle x : [61; 99]\rangle\}$$
$$N_2 = \{\langle x : (-\infty; -1]\rangle, \langle x : [61; 99]\rangle\} \qquad N_6 = \{\langle x : (-\infty; 0]\rangle, \langle x : [61; 100]\rangle\}$$
$$N_3 = \bot \qquad N_7 = \top$$
$$N_5 = \{\langle x : (-\infty; -1]\rangle, \langle x : [61; 99]\rangle\} \qquad N_{10} = \top$$

To initialize the positive side for the loop, we use a universal recurrent set obtained by three calls to E-HSF with different recurrent set templates. The result is $R_\forall = \{\langle x : [4; 60]\rangle, \langle x : [100; +\infty)\rangle\}$. Note that in this example, universal recurrence and safety coincide, and our analysis will be able to improve the result by showing that the states in $\langle x : [1; 3]\rangle$ are also safe. Since we are using a power set domain, we choose to use *pre* instead of *wp* for the final statement (as described in Sect. 4.1), not just for the loop (where we need to use it due to starting with $R_\forall$). We start with

$$P_{10}^1 = \mathit{pre}(\texttt{assert}\,0, \top) - N_{10} = \bot - N_{10} = \bot$$

then proceed to the loop (again, computing *pre* of its body in steps),

$$P_1^1 = R_\forall \sqcup [x \leq 0, P_{10}^1]^\sharp - N_1 = \{\langle x : [4; 60]\rangle, \langle x : [100; +\infty)\rangle\}$$
$$P_7^1 = \mathit{pre}(x \leftarrow 0, P_1^1) - N_7 = \bot$$
$$\begin{aligned} P_6^1 &= [x = 100, P_7^1]^\sharp \sqcup [x \neq 100, P_1^1]^\sharp - N_6 \\ &= \{\langle x : [4 : 60]\rangle, \langle x : [101; +\infty)\rangle\} - N_6 \\ &= \{\langle x : [4 : 60]\rangle, \langle x : [101; +\infty)\rangle\} \end{aligned}$$
$$P_5^1 = \mathit{pre}(x \leftarrow x + 1, P_6^1) - N_5 = \{\langle x : [3 : 59]\rangle, \langle x : [100; +\infty)\rangle\}$$
$$P_3^1 = \mathit{pre}(x \leftarrow 50, P_5^1) - N_3 = \top$$
$$\begin{aligned} P_2^1 &= [x = 60, P_3^1]^\sharp \sqcup [x \neq 60, P_5^1]^\sharp - N_2 \\ &= \{\langle x : [3; 59]\rangle, \langle x : [60]\rangle, \langle x : [100; +\infty)\rangle\} - N_2 \\ &= \{\langle x : [3; 60]\rangle, \langle x : [100; +\infty)\rangle\} \end{aligned}$$
$$P_1^2 = (P_1^1 \sqcup ([x \geq 1, P_2^1]^\sharp - N_2)) - N_2 = \{\langle x : [3; 60]\rangle, \langle x : [100; +\infty)\rangle\}$$

at which point we detect an unstable bound, but we again proceed without widening and are able to discover the stable bound of 1. Also note that (as observed in Sect. 4.1),

P_1 is bounded by $P_{10} \sqcup [\neg x \leq 0]^\sharp = \langle x : [1; +\infty)\rangle$. This bound could be used to improve the result of widening. Thus, we take

$$
\begin{array}{ll}
P_1 = \{\langle x : [1; 60]\rangle, \langle x : [100; +\infty)]\rangle\} & \\
P_2 = \{\langle x : [0; 60]\rangle, \langle x : [100; +\infty)]\rangle\} & P_6 = \{\langle x : [1; 60]\rangle, \langle x : [101; +\infty)]\rangle\} \\
P_3 = \top & P_7 = \bot \\
P_5 = \{\langle x : [0; 59]\rangle, \langle x : [100; +\infty)]\rangle\} & P_{10} = \bot
\end{array}
$$

Thus, in this example, our analysis was able to prove that initial states $\{\langle x : [1; 60]\rangle, \langle x : [100; +\infty)]\rangle\}$ are safe, which is a slight improvement over the output of E-HSF.

Example 2. In this example, we consider the program in Fig. 3. In the program, $*$ stands for a value non-deterministically chosen at runtime. All the assumptions made for Example 1 are in effect for this one as well, except that we increase the bound on the size of the domain element to 4. The analysis is able to produce the following approximation of the safe entry states:

$$
\begin{array}{c}
\{\langle x : [100; +\infty), y : \top\rangle, \langle x : (-\infty; 0], y : (-\infty; -1]\rangle, \\
\langle x : (-\infty; 0], y : [1; +\infty)\rangle, \langle x : [1; 99], y : [1; +\infty)\rangle\}
\end{array}
$$

This example also displays an interplay between coarse subtraction and the use of over-approximate operations (especially, *pre*) on the positive side. In order to retain precision when coarse subtraction is used, it seems important to be able to keep the positive side partitioned into a number of disjuncts. In a real-world analysis, this can be achieved, e.g., by some form of trace partitioning [15]. In this example, we employ a few simple tricks, one of those can be seen from Fig. 3. Observe that we translated the non-deterministic condition in lines 3-7 of the syntactically sugared program (Fig. 3a) into equivalent nested conditions (statement 3 of the desugared program in Fig. 3b) which allows the necessary disjuncts to emerge on the positive side.

Shape Analysis Examples In what follows, we demonstrate our approach for a shape analysis domain. We treat two simple examples using the domain of 3-valued structures, and we claim that our approach provides a viable decomposition of backward analysis (for this domain and probably for some other shape analysis domains). For background information on shape analysis with 3-valued logic, please refer to [21] and accompanying papers, e.g., [18,1,24]. To handle the examples, we use a mechanized procedure built on top of the TVLA shape analysis tool (`http://www.cs.tau.ac.il/~tvla/`).

Example 3. In this example, we consider the program in Fig. 4. The program manipulates a pointer variable x, and the heap cells each have a pointer field n. We compare x to *nil* to check whether it points to a heap cell. We write $x \to n$ to denote access to the pointer field n of the heap cell pointed to by x. The program in Fig. 4 just traverses its input structure in a loop.

The analysis identifies that both cyclic and acyclic lists are safe inputs for the program – and summarizes them in eight and nine shapes respectively. Figures 6 and 7 show examples of the resulting shapes.

```
1  while x ≥ 1 do
2    if x ≤ 99 then
3      if y ≤ 0 ∧ ∗ then
4
5        assert 0
6
7      end
8      if ∗ then
9        x ← −1
10     end
11   end
12   x ← x + 1
13 end
14 assert y ≠ 0
```

(a) With syntactic sugar.

$$
\begin{aligned}
&{}_1([x \geq 1];\\
&\quad {}_2(([x \leq 99];\\
&\qquad {}_3(([y \leq 0]; {}_4({}_5\,\texttt{assert}\,0 + \texttt{skip}))\\
&\qquad\quad + ([y \geq 1]; \texttt{skip}));\\
&\qquad {}_8({}_9 x \leftarrow -1 + \texttt{skip})\\
&\quad) + ([x \geq 100]; \texttt{skip}));\\
&\quad {}_{12} x \leftarrow x + 1\\
&)^*; [x \leq 0];\\
&{}_{14}\,\texttt{assert}\, y \neq 0
\end{aligned}
$$

(b) Desugared.

Fig. 3: Example program 2.

```
1  while x ≠ nil do
2    x ← (x → n)
3  end
```

Fig. 4: Example program 3.

```
1  while x ≠ nil do
2    x ← (x → n)
3    x ← (x → n)
4  end
```

Fig. 5: Example program 4.

Example 4. In this example, we consider the program in Fig. 5. In this program, the loop body makes two steps through the list instead of just one. While the first step (at line 2) is still guarded by the loop condition, the second step (at line 3) is a source of failure. That is, the program fails when given a list of odd length as an input. The abstraction that we employ is not expressive enough to encode such constraints on the length of the list. The analysis is able to show that cyclic lists represent safe inputs, but the only acyclic list that the analysis identifies as safe is the list of length exactly two.

6 Related Work

In [14], a backward shape analysis with 3-valued logic is presented that relies on the correspondence between 3-valued structures and first-order formulas [24]. It finds an over-approximation of states that may lead to failure, and then (as 3-valued structures do not readily support complementation) the structures are translated to an equivalent quantified first-order formula, which is then negated. This corresponds to approximating the negative side in our approach and then taking the complement, with the exception that the result is not represented as an element of the abstract domain. At least in principle, the symbolic abstraction $\hat{\alpha}$ of [19] could map back to the abstract domain.

For shape analysis with separation logic [20], preconditions can be inferred using a form of abduction called bi-abduction [5]. The analysis uses an over-approximate

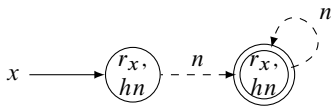


Fig. 6: Example of a safe structure causing non-termination.

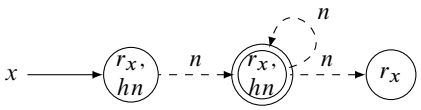


Fig. 7: Example of a safe structure leading to successful termination.

abstraction, and it includes a filtering step that checks generated preconditions (by computing their respective postconditions) and discards the unsound ones. The purpose of the filtering step – keeping soundness of a precondition produced with over-approximate abstraction – is similar to our use of the negative side.

For numeric programs, the problem of finding preconditions for safety has seen some attention lately. In [17], a numeric analysis is presented that is based primarily on over-approximation. It simultaneously computes the representations of two sets: of states that may lead to successful termination, and of states that may lead to failure. Then, meet and generic negation are used to produce representations of states that cannot fail, states that must fail, etc. An under-approximating backward analysis for the polyhedral domain is presented in [16]. The analysis defines the appropriate under-approximate abstract transformers and to ensure termination, proposes a lower widening based on the generator representation of polyhedra. With E-HSF [4], the search for preconditions can be formulated as solving $\forall\exists$ quantified Horn clauses extended with well-foundedness conditions. The analysis is targeted specifically at linear programs, and is backed by a form of counterexample-guided abstraction refinement.

7 Conclusion and Future Work

We presented an alternative decomposition of backward analysis, suitable for domains that do not readily support complementation and under-approximation of greatest fixed points. Our approach relies on an under-approximating subtraction operation and a procedure that finds recurrent sets for loops – and builds a sequence of successive under-approximations of the safe states. This decomposition allowed us to implement a backwards analysis for the domain of 3-valued structures and to obtain acceptable analysis results for two simple programs.

For shape analysis examples, we employed quite a simplistic procedure to approximate a recurrent set. One direction for future research is into recurrence search procedures for shape analysis that are applicable to realistic programs.

Another possible direction is to explore the settings where non-termination counts as failure. This is the case, e.g., when checking abstract counterexamples for concrete feasibility [3].

Acknowledgements We thank Andrey Rybalchenko for helpful discussion and assistance with E-HSF, and Mooly Sagiv and Roman Manevich for sharing the source code of TVLA. A. Bakhirkin is supported by a Microsoft Research PhD Scholarship.

References

1. Arnold, G., Manevich, R., Sagiv, M., Shaham, R.: Combining shape analyses by intersecting abstractions. In: Emerson, E.A., Namjoshi, K.S. (eds.) VMCAI. LNCS, vol. 3855, pp. 33–48. Springer (2006)
2. Bagnara, R., Hill, P.M., Zaffanella, E.: Widening operators for powerset domains. STTT 9(3-4), 413–414 (2007)
3. Berdine, J., Bjørner, N., Ishtiaq, S., Kriener, J.E., Wintersteiger, C.M.: Resourceful reachability as HORN-LA. In: McMillan, K.L., Middeldorp, A., Voronkov, A. (eds.) LPAR. LNCS, vol. 8312, pp. 137–146. Springer (2013)
4. Beyene, T.A., Popeea, C., Rybalchenko, A.: Solving existentially quantified Horn clauses. In: Sharygina, N., Veith, H. (eds.) CAV. LNCS, vol. 8044, pp. 869–882. Springer (2013)
5. Calcagno, C., Distefano, D., O'Hearn, P.W., Yang, H.: Compositional shape analysis by means of bi-abduction. In: Shao, Z., Pierce, B.C. (eds.) POPL. pp. 289–300. ACM (2009)
6. Calcagno, C., Ishtiaq, S.S., O'Hearn, P.W.: Semantic analysis of pointer aliasing, allocation and disposal in Hoare logic. In: PPDP. pp. 190–201 (2000)
7. Calcagno, C., Yang, H., O'Hearn, P.W.: Computability and complexity results for a spatial assertion language for data structures. In: APLAS. pp. 289–300 (2001)
8. Clarke, E.M.: Program invariants as fixed points (preliminary reports). In: FOCS. pp. 18–29. IEEE Computer Society (1977)
9. Cousot, P.: Semantic foundations of program analysis. In: Muchnick, S.S., Jones, N.D. (eds.) Program Flow Analysis: Theory and Applications, pp. 303–342. Prentice-Hall (1981)
10. Cousot, P., Cousot, R.: Abstract interpretation and application to logic programs. J. Log. Program. 13(2&3), 103–179 (1992)
11. Cousot, P., Halbwachs, N.: Automatic discovery of linear restraints among variables of a program. In: Aho, A.V., Zilles, S.N., Szymanski, T.G. (eds.) POPL. pp. 84–96. ACM Press (1978)
12. Gupta, A., Henzinger, T.A., Majumdar, R., Rybalchenko, A., Xu, R.G.: Proving non-termination. In: Necula, G.C., Wadler, P. (eds.) POPL. pp. 147–158. ACM (2008)
13. Halbwachs, N., Proy, Y.E., Roumanoff, P.: Verification of real-time systems using linear relation analysis. Form. Method. Syst. Des. 11(2), 157–185 (1997)
14. Lev-Ami, T., Sagiv, M., Reps, T., Gulwani, S.: Backward analysis for inferring quantified preconditions. Tech. Rep. TR-2007-12-01, Tel Aviv University (Dec 2007)
15. Mauborgne, L., Rival, X.: Trace partitioning in abstract interpretation based static analyzers. In: Sagiv, S. (ed.) ESOP. LNCS, vol. 3444, pp. 5–20. Springer (2005)
16. Miné, A.: Inferring sufficient conditions with backward polyhedral under-approximations. Electr. Notes Theor. Comput. Sci. 287, 89–100 (2012)
17. Popeea, C., Chin, W.N.: Dual analysis for proving safety and finding bugs. Sci. Comput. Program. 78(4), 390–411 (2013)
18. Reps, T.W., Sagiv, S., Loginov, A.: Finite differencing of logical formulas for static analysis. In: Degano, P. (ed.) ESOP. LNCS, vol. 2618, pp. 380–398. Springer (2003)
19. Reps, T.W., Sagiv, S., Yorsh, G.: Symbolic implementation of the best transformer. In: Steffen, B., Levi, G. (eds.) VMCAI. LNCS, vol. 2937, pp. 252–266. Springer (2004)
20. Reynolds, J.C.: Separation logic: A logic for shared mutable data structures. In: LICS. pp. 55–74. IEEE Computer Society (2002)
21. Sagiv, S., Reps, T.W., Wilhelm, R.: Parametric shape analysis via 3-valued logic. ACM Trans. Program. Lang. Syst. 24(3), 217–298 (2002)
22. Simon, A., King, A.: Widening polyhedra with landmarks. In: Kobayashi, N. (ed.) APLAS. LNCS, vol. 4279, pp. 166–182. Springer (2006)

23. Velroyen, H., Rümmer, P.: Non-termination checking for imperative programs. In: Beckert, B., Hähnle, R. (eds.) TAP. LNCS, vol. 4966, pp. 154–170. Springer (2008)
24. Yorsh, G., Reps, T.W., Sagiv, M., Wilhelm, R.: Logical characterizations of heap abstractions. ACM Trans. Comput. Log. 8(1) (2007)

A Proofs

Lemma 1. *For a statement C and a set of states $S \subseteq \mathcal{U}$, $wp(C, S) = \mathcal{U} \smallsetminus pre\text{+}fail(C, \mathcal{U} \smallsetminus S)$.*

Proof.

$$\begin{aligned}
&\mathcal{U} \smallsetminus \big(pre\text{+}fail(C, \mathcal{U} \smallsetminus S)\big) \\
&= \mathcal{U} \smallsetminus \big(pre(C, \mathcal{U} \smallsetminus S) \cup fail(C)\big) \\
&= \mathcal{U} \smallsetminus (\{s \in \mathcal{U} \mid \exists s' \in \mathcal{U} \smallsetminus S.\, [\![C]\!](s, s')\} \cup \{s \in \mathcal{U} \mid [\![C]\!](s, \epsilon)\}) \\
&= \mathcal{U} \smallsetminus \{s \in \mathcal{U} \mid \exists s' \in (\mathcal{U} \smallsetminus S) \cup \{\epsilon\}.\, [\![C]\!](s, s')\} \\
&= \{s \in \mathcal{U} \mid \nexists s' \in (\mathcal{U} \smallsetminus S) \cup \{\epsilon\}.\, [\![C]\!](s, s')\} \\
&= \{s \in \mathcal{U} \mid \forall s' \in (\mathcal{U} \smallsetminus S) \cup \{\epsilon\}.\, \neg[\![C]\!](s, s')\} \\
&= \{s \in \mathcal{U} \mid \forall s' \in \mathcal{U} \cup \{\epsilon\}.\, [\![C]\!](s, s') \Rightarrow s' \in S\} \\
&= wp(C, S)
\end{aligned}$$

□

Lemma 2. *For a statement C and set of states $S \subseteq \mathcal{U}$, $P(C, S) = \mathcal{U} \smallsetminus N(C, \mathcal{U} \smallsetminus S)$.*

Proof (Sketch). Proceed by induction on the structure of C. For atomic statements the result follows from Lemma 1. For sequential composition and branch, the result follows directly from the induction hypothesis.

It remains to consider loops C^*. Let A be a fixed point of $\lambda X.\, P(C, X) \cap S$, i.e., $A = P(C, A) \cap S$. Let $B = \mathcal{U} \smallsetminus A$:

$$\begin{aligned}
B &= \mathcal{U} \smallsetminus \big(P(C, \mathcal{U} \smallsetminus B) \cap S\big) \\
&= \big(\mathcal{U} \smallsetminus P(C, \mathcal{U} \smallsetminus B)\big) \cup (\mathcal{U} \smallsetminus S)
\end{aligned}$$

by structural induction hypothesis

$$= N(C, B) \cup (\mathcal{U} \smallsetminus S)$$

That is, B is a fixed point of $\lambda Y.\, N(C, Y) \cup (\mathcal{U} \smallsetminus S)$.

A similar argument shows that if B is a fixed point of $\lambda Y.\, N(C, Y) \cup (\mathcal{U} \smallsetminus S)$, then $\mathcal{U} \smallsetminus B$ is a fixed point of $\lambda X.\, P(C, X) \cap S$.

Let $A' = P(C^*, S) = \operatorname{gfp} \lambda X.\, P(C, X) \cap S$, hence $\mathcal{U} \smallsetminus A'$ is a fixed point of $\lambda Y.\, N(C, Y) \cup (\mathcal{U} \smallsetminus S)$. Let $B' = N(C^*, \mathcal{U} \smallsetminus S) = \operatorname{lfp} \lambda Y.\, N(C, Y) \cup (\mathcal{U} \smallsetminus S)$, hence $\mathcal{U} \smallsetminus B'$ is a fixed point of $\lambda X.\, P(C, X) \cap S$. Since A' is maximal, $A' \supseteq \mathcal{U} \smallsetminus B'$. Since B' is minimal, $B' \subseteq \mathcal{U} \smallsetminus A'$, and $A' \subseteq \mathcal{U} \smallsetminus B'$. Hence, $A' = \mathcal{U} \smallsetminus B'$, i.e., $P(C^*, S) = \mathcal{U} \smallsetminus N(C^*, \mathcal{U} \smallsetminus S)$. □

Lemma 3. *For a statement C and sets of states $S, V \subseteq \mathcal{U}$, $P(C, S) = wp(C, S)$, and $N(C, V) = pre\text{+}fail(C, V)$.*

Proof (Sketch). It is enough to prove the Lemma for one (e.g., negative) side, then for the other side, it follows from Lemmas 1 and 2.

For the negative side, the proof proceeds by structural induction. For the atomic statements, sequential composition and branching, it follows from the definitions of the semantics and *pre+fail*.

Let the composition of relations in $\mathcal{U} \times (\mathcal{U} \cup \{\epsilon\})$ be lifted to iterated relational composition as follows: $R^0 = \Delta_{\mathcal{U}}$, and $R^{i+1} = R \fatsemi R^i$.

We also lift *pre+fail* from statements to relations. For $R \subseteq \mathcal{U} \times (\mathcal{U} \cup \{\epsilon\})$ and $V \subseteq \mathcal{U}$,

$$\mathit{pre+fail}'(R, V) = \{s \in \mathcal{U} \mid \exists s' \in V \cup \{\epsilon\}.\ R(s, s')\}$$

Note that for a statement C and set of states $V \subseteq \mathcal{U}$,

$$\mathit{pre+fail}(C, V) = \mathit{pre+fail}'([\![C]\!], V) \tag{6}$$

Note that from the structural induction hypothesis,

$$\mathit{pre+fail}'([\![C_1]\!] \fatsemi [\![C_2]\!], V) = \mathit{pre+fail}'([\![C_1]\!], \mathit{pre+fail}'([\![C_2]\!], V))$$

and therefore, for any i, we can show by induction on i,

$$\mathit{pre+fail}'([\![C]\!]^i, V) = (\lambda X.\ \mathit{pre+fail}'([\![C]\!], X))^i\ V \tag{7}$$

For a loop, C^* and a set of states $V \subseteq \mathcal{U}$,

$N(C^*, V)$

by definition

$= \mathrm{lfp}\, \lambda Y.\ N(C, Y) \cup V$

by structural induction hypothesis and (6)

$= \mathrm{lfp}\, \lambda Y.\ \mathit{pre+fail}'([\![C]\!], Y) \cup V$

by Kleene's Fixed Point Theorem

$= \bigcup_{i=0}^{\infty} (\lambda Y.\ \mathit{pre+fail}'([\![C]\!], Y) \cup V)^i\ \varnothing$

by continuity of union

$= \bigcup_{i=0}^{\infty} (\lambda Y.\ \mathit{pre+fail}'([\![C]\!], Y))^i\ V$

by (7)

$= \bigcup_{i=0}^{\infty} \mathit{pre+fail}'([\![C]\!]^i, V)$

by continuity of $\mathit{pre+fail}'$ in its first argument

$= \mathit{pre+fail}'(\bigcup_{i=0}^{\infty} [\![C]\!]^i, V)$

by Kleene's Fixed Point Theorem

$= \mathit{pre+fail}'(\mathrm{lfp}\, \lambda Y.\ \Delta_{\mathcal{U}} \cup ([\![C]\!] \fatsemi Y), V)$

by definition of $[\![C^*]\!]$ and (6)

$= \mathit{pre+fail}([\![C^*]\!], V)$

□

Lemma 4. *For the loop* $C_{\text{loop}} = ([\psi]\ ;\ C_{\text{body}})^*\ ;\ [\varphi]$*, and a set of states* $S \subseteq \mathcal{U}$

$$R_\forall \subseteq P(C_{\text{loop}}, S) \qquad\qquad R_\exists \smallsetminus N(C_{\text{loop}}, \mathcal{U} \smallsetminus S) \subseteq P(C_{\text{loop}}, S)$$

Proof. The result for the existential recurrent set $R_\exists$ follows from Lemma 2 independently of the definition of $R_\exists$. We proceed to prove the result for the universal recurrent set.

Let $R = P(C_{\text{loop}}, \varnothing)$, then as in (3),

$$R = \operatorname{gfp} \lambda X.\ \big([\![\neg\psi]\!] \cup P(C_{\text{body}}, X)\big) \cap [\![\neg\varphi]\!]$$

using $[\![\neg\varphi]\!] \cap [\![\neg\psi]\!] = \varnothing$, and Lemma 3

$$= \operatorname{gfp} \lambda X.\ wp(C_{\text{body}}, X) \cap [\![\neg\varphi]\!]$$

Since P is monotone in its second argument (as wp is), then for every S it holds that $R \subseteq P(C_{\text{loop}}, S)$. From definition of universal recurrent set,

$$R_\forall \subseteq [\![\neg\varphi]\!]$$
$$\forall s \in R_\forall.\ \big(\forall s' \in \mathcal{U} \cup \{\epsilon\}.\ [\![C_{\text{body}}]\!](s, s') \Rightarrow s' \in R_\forall\big)$$

From definition of wp

$$wp(C_{\text{body}}, R_\forall) = \{s \in \mathcal{U} \mid \forall s' \in \mathcal{U} \cup \{\epsilon\}.\ [\![C_{\text{body}}]\!](s, s') \Rightarrow s' \in R_\forall\}$$

Hence

$$R_\forall \subseteq [\![\neg\varphi]\!] \text{ and } R_\forall \subseteq wp(C_{\text{body}}, R_\forall)$$
$$R_\forall \subseteq wp(C_{\text{body}}, R_\forall) \cap [\![\neg\varphi]\!]$$

From Tarski's Fixed Point Theorem

$$R_\forall \subseteq \operatorname{gfp} \lambda X.\ wp(C_{\text{body}}, X) \cap [\![\neg\varphi]\!]$$

That is, for every S

$$R_\forall \subseteq R \subseteq P(C_{\text{loop}}, S)$$

□

Theorem 1. *The alternative characterizations of the positive side of the loop:* (4) *and* (5) *– under-approximate the original characterization* (3)*. That is, for a set* $S \subseteq \mathcal{U}$*,*

$$P^\exists(C_{\text{loop}}, S) \subseteq P(C_{\text{loop}}, S) \qquad\qquad P^\forall(C_{\text{loop}}, S) \subseteq P(C_{\text{loop}}, S)$$

Proof. Let $C_{\text{loop}} = ([\psi]\ ;\ C_{\text{body}})^*\ ;\ [\varphi]$ be a loop as in (2), $R_\exists$ be its existential recurrent set, P_{rest} be a set of states, $N_{\text{rest}} = \mathcal{U} \smallsetminus P_{\text{rest}}$, $P_{\text{loop}} = P(C_{\text{loop}}, P_{\text{rest}})$, $N_{\text{loop}} = N(C_{\text{loop}}, N_{\text{rest}}) = U \smallsetminus P_{\text{loop}}$ (i.e., P_{loop} and N_{loop} are the original characterizations of the positive and negative sides as in (3)). Then, it holds that

$$[\![\neg\psi]\!] \cap P_{\text{rest}} \subseteq P_{\text{loop}} \tag{8}$$
$$R_\exists \smallsetminus N_{\text{loop}} \subseteq P_{\text{loop}} \tag{9}$$
$$\text{For } S \subseteq P_{\text{loop}} \text{ we have } wp(C_{\text{body}}, S) \cap ([\![\neg\varphi]\!] \cup ([\![\psi]\!] \cap P_{\text{rest}})) \subseteq P_{\text{loop}} \tag{10}$$

Equation (8) can be seen from (3) and describes the states that immediately cause successful termination of the loop. Equation (9) is due to Lemma 4. Equation (10) is due to the following.

$$P_{\text{loop}} = \text{gfp}\,\lambda X.\,\big([\![\neg\psi]\!] \cup P(C_{\text{body}}, X)\big) \cap ([\![\neg\varphi]\!] \cup P_{\text{rest}})$$

Then

$$P_{\text{loop}} = \big([\![\neg\psi]\!] \cup P(C_{\text{body}}, P_{\text{loop}})\big) \cap ([\![\neg\varphi]\!] \cup P_{\text{rest}})$$

If $S \subseteq P_{\text{loop}}$ then $([\![\neg\psi]\!] \cup wp(C_{\text{body}}, S)) \cap ([\![\neg\varphi]\!] \cup P_{\text{rest}}) \subseteq P_{\text{loop}}$. Also,

$$([\![\neg\psi]\!] \cup wp(C_{\text{body}}, S)) \cap ([\![\neg\varphi]\!] \cup P_{\text{rest}})$$

Due to $[\![\neg\varphi]\!] \cap [\![\neg\psi]\!] = \varnothing$

$$= ([\![\neg\psi]\!] \cup P_{\text{rest}}) \cup (wp(C_{\text{body}}, S) \cap ([\![\neg\varphi]\!] \cup P_{\text{rest}}))$$

splitting the second occurrence of $P_{\text{rest}} = ([\![\psi]\!] \cap P_{\text{rest}}) \cup ([\![\neg\psi]\!] \cap P_{\text{rest}})$

$$\begin{aligned} &= ([\![\neg\psi]\!] \cap P_{\text{rest}}) \cup \\ &\quad (wp(C_{\text{body}}, S) \cap ([\![\neg\varphi]\!] \cup ([\![\psi]\!] \cap P_{\text{rest}}))) \cup \\ &\quad (wp(C_{\text{body}}, S) \cap [\![\neg\psi]\!] \cap P_{\text{rest}}) \end{aligned}$$

Note that the first and last disjuncts are known to be included P_{loop} due to (8), and we prefer not to count them here. Thus, if $S \subseteq P_{\text{loop}}$, then $wp(C_{\text{body}}, S) \cap ([\![\neg\varphi]\!] \cup ([\![\psi]\!] \cap P_{\text{rest}})) \subseteq P_{\text{loop}}$.

Let $P^{\exists}_{\text{loop}} = P^{\exists}(C_{\text{loop}}, P_{\text{rest}})$. Using (4), we characterize $P^{\exists}_{\text{loop}}$ as the limit of the ascending (due to monotonicity of *wp*) chain:

$$P^{\exists}_{\text{loop},0} \subseteq P^{\exists}_{\text{loop},1} \subseteq P^{\exists}_{\text{loop},2} \subseteq \ldots$$

where

$$\begin{aligned} P^{\exists}_{\text{loop},0} &= ([\![\neg\psi]\!] \cap P_{\text{rest}}) \cup (R_{\exists} \smallsetminus N_{\text{loop}}) \\ P^{\exists}_{\text{loop,k}+1} &= P^{\exists}_{\text{loop},0} \cup \Big(wp(C_{\text{body}}, P^{\exists}_{\text{loop,k}}) \cap \big([\![\neg\varphi]\!] \cup ([\![\psi]\!] \cap P_{\text{rest}})\big)\Big) \end{aligned}$$

and $P^{\exists}_{\text{loop}} = \bigcup_{k=0}^{\infty} P^{\exists}_{\text{loop,k}}$. From (8), (9), and (10), it follows that for each element of the chain, $P^{\exists}_{\text{loop,k}} \subseteq P_{\text{loop}}$, and $P^{\exists}_{\text{loop}} \subseteq P_{\text{loop}}$.

For $P^{\forall}_{\text{loop}}$, the proof proceeds in a similar way. For a universal recurrent set $R_{\forall}$ and for the solution of the original equations (3), it holds that

$$([\![\varphi]\!] \cap P_{\text{rest}}) \smallsetminus N_{\text{loop}} \subseteq P_{\text{loop}} \tag{11}$$

$$R_{\forall} \subseteq P_{\text{loop}} \tag{12}$$

$$\text{For } S \subseteq P_{\text{loop}} \text{ we have } ([\![\neg\varphi]\!] \cap pre(C_{\text{body}}, S)) \smallsetminus N_{\text{loop}} \subseteq P_{\text{loop}} \tag{13}$$

Equations (11) and (13) are due to Lemma 2. Equation (12) is due to Lemma 4. Using (5), we characterize $P^{\forall}_{\text{loop}}$, as the limit of the ascending (due to monotonicity of *pre*) chain: $P^{\forall}_{\text{loop}} = \bigcup_{k=0}^{\infty} P^{\forall}_{\text{loop,k}}$, where

$$P^{\forall}_{\text{loop},0} = R_{\forall} \cup \big(([\![\varphi]\!] \cap P_{\text{rest}}) \smallsetminus N_{\text{loop}}\big)$$

$$P^{\forall}_{\text{loop,k+1}} = P^{\forall}_{\text{loop},0} \cup \left(\left([\![\neg\varphi]\!] \cap \mathit{pre}(C_{\text{body}}, P^{\forall}_{\text{loop,k}}) \right) \smallsetminus N_{\text{loop}} \right)$$

From (11), (12), and (13) it follows that for all $k \geq 0$, $P^{\forall}_{\text{loop,k}} \subseteq P_{\text{loop}}$, and $P^{\forall}_{\text{loop}} \subseteq P_{\text{loop}}$. □

Theorem 2. *For a statement C and $p, n \in \mathcal{D}$ s.t. $\gamma(p) \cap \gamma(n) = \varnothing$, $N^\sharp(C, n) \supseteq N(C, \gamma(n))$ and $P^\flat(C, p, n) \subseteq P(C, \mathcal{U} \smallsetminus \gamma(n))$. Hence, for a top-level program C_{prg}, $\gamma(N^\sharp(C_{\text{prg}}, \bot)) \supseteq N(C_{\text{prg}}, \varnothing)$ (i.e., it over-approximates input states that may lead to failure), and $\gamma(P^\flat(C_{\text{prg}}, \top, \bot)) \subseteq P(C_{\text{prg}}, \mathcal{U})$ (i.e., it under-approximates safe input states).*

Proof (Sketch). For a loop-free statement, the result for $N^\sharp$ can be seen by induction over its structure: it follows from monotonicity of (concrete) N in its second argument, and from that the abstract transformers used in the equations for $N^\sharp$ are over-approximating. For a loop $C_{\text{loop}} = ([\psi]; C_{\text{body}})^*; [\varphi]$, we can abbreviate the characterizations using a pair of functions F and $F^\sharp$, s.t., F is monotone and

$$N(C_{\text{loop}}, S) = \operatorname{lfp} \lambda Y.\, F(Y, S) = \operatorname{lfp} \lambda Y.\, ([\![\varphi]\!] \cap S) \cup F(Y, S)$$
$$N^\sharp(C_{\text{loop}}, n) = \text{the first } n_j \in \{n_i\}_{i \geq 0} \text{ such that } n_{j+1} \sqsubseteq n_j \text{ where}$$
$$n_0 = [\varphi, n]^\sharp \text{ and } n_{i+1} = n_i \mathbin{\triangledown} F^\sharp(n_i, n)$$

Using the structural induction hypothesis, we can show that $F^\sharp$ over-approximates F, i.e.,

$$\gamma(F^\sharp(Y, n)) \supseteq F(\gamma(Y), \gamma(n))$$

From the stopping condition of the ascending chain of approximations of $N^\sharp(C_{\text{loop}}, n)$:

$$N^\sharp(C_{\text{loop}}, n) \sqsupseteq N^\sharp(C_{\text{loop}}, n) \mathbin{\triangledown} F^\sharp(N^\sharp(C_{\text{loop}}, n), n)$$

Since γ is monotone, and $\triangledown$ over-approximates $\cup$,

$$\gamma(N^\sharp(C_{\text{loop}}, n)) \supseteq \gamma(N^\sharp(C_{\text{loop}}, n)) \cup \gamma(F^\sharp(N^\sharp(C_{\text{loop}}, n), n))$$

Since $F^\sharp$ over-approximates F and $[\varphi, \cdot]^\sharp$ over-approximates $[\![\varphi]\!] \cap \gamma(\cdot)$,

$$\gamma(N^\sharp(C_{\text{loop}}, n)) \supseteq ([\![\varphi]\!] \cap \gamma(n)) \cup \gamma(N^\sharp(C_{\text{loop}}, n)) \cup F(\gamma(N^\sharp(C_{\text{loop}}, n)), \gamma(n))$$

From monotonicity of F and Tarski's Fixed Point Theorem,

$$\begin{aligned} \gamma(N^\sharp(C_{\text{loop}}, n)) &\supseteq \operatorname{lfp} \lambda Y.\, ([\![\varphi]\!] \cap \gamma(n)) \cup Y \cup F(Y, \gamma(n)) \\ &= \operatorname{lfp} \lambda Y.\, ([\![\varphi]\!] \cap \gamma(n)) \cup F(Y, \gamma(n)) \\ \gamma(N^\sharp(C_{\text{loop}}, n)) &\supseteq N(C_{\text{loop}}, \gamma(n)) \end{aligned}$$

Exactly as we wanted to show.

For $P^\flat$, the result follows from the use of subtraction. For a statement C and disjoint $d, n \in \mathcal{D}$, $P^\flat(C, d, n)$ is defined as $q - N^\sharp(C, n)$ for some $q \in \mathcal{D}$. From the definition of subtraction,

$$\gamma(P^\flat(C, d, n)) \subseteq \mathcal{U} \smallsetminus \gamma(N^\sharp(C, n))$$

From the result for $N^\sharp$

$$\gamma(P^\flat(C,d,n)) \subseteq \mathcal{U} \smallsetminus N(C,\gamma(n))$$
$$\gamma(P^\flat(C,d,n)) \subseteq P(C,\mathcal{U} \smallsetminus \gamma(n))$$

Then, for a top-level program C_{prg},

$$\gamma(P^\flat(C,\top,\bot)) \subseteq P(C_{\text{prg}},\mathcal{U})$$

□

B Extended examples

Example 1 (extended) Input to E-HSF is shown in Fig. 8.

```
1  next(X0, X3) :=
2    X0>0,
3    (X0=600 -> X1=50 ; X1=X0),
4    X2=X1+1,
5    (X2=1000 -> X3=0 ; X3=X2).
6
7  % One template is uncommented per query.
8  % Gives [4; 60].
9  skolem_template(s1,[X],true,true,(A=<X, X=<B),(A=<B)).
10 % Gives [100; +inf).
11 % skolem_template(s1,[X],true,true,(A=<X),true).
12 % Gives nothing.
13 % skolem_template(s1,[X],true,true,(X=<B),true).
14
15 exists([X], rec(X)).
16 X>=1 :- rec(X).
17 rec(X1) :- rec(X0), next(X0,X1).
```

Fig. 8: Input to E-HSF to find the universal recurrent set for Example 1.

Example 2 (extended) Again, we start with the negative side. For the final location,

$$N_{14}^1 = \textit{pre+fail}(\texttt{assert}\ y \neq 0, \bot) = \langle x : \top, y : [0] \rangle$$

then, proceed to the loop

$$N_1^1 = [x \leq 0, N_{14}^1]^\natural = \langle x : (-\infty, 0], y : [0] \rangle$$
$$N_{12}^1 = \textit{pre+fail}(x \leftarrow x+1, N_1^1) = \langle x : (-\infty; -1], y : [0] \rangle$$
$$N_9^1 = \textit{pre+fail}(x \leftarrow -1, N_{12}^1) = \langle x : \top, y : [0] \rangle$$
$$N_8^1 = N_9^1 \sqcup N_{12}^1 = \langle x : \top, y : [0] \rangle$$

```
1  next(X0, Y0, E0, X2, Y2, E2) :=
2    X0 > 0,
3    E0 = 0,
4    (X0<100 ->
5       (((Y0=<0, E1=1);(E1=E0)),
6        ((X1=(-1));(X1=X0)));
7      (X1=X0, E1=E0)),
8    X2 = X1+1,
9    Y2 = Y0,
10   E2 = E1.
11
12 % Gives [100; +inf).
13 skolem_template(s1,[X,Y,E],true,true,(A=<X),true).
14
15 (X>=1, E==0) :- rec(X, Y, E).
16 rec(X1,Y1,E1) :- rec(X0,Y0,E0), next(X0,Y0,E0,X1,Y1,E1).
```

Fig. 9: Input to E-HSF to find the universal recurrent set for Example 2.

$$N_5^1 = \top$$
$$N_3^1 = [y \le 0, N_5^1]^\natural \sqcup N_8^1 = \langle x : \top, y : (-\infty; 0]\rangle$$
$$N_2^1 = [x \le 99, N_3^1]^\natural \sqcup [x \ge 100, N_{12}^1]^\natural = \langle x : (-\infty; 99], y : (-\infty; 0]\rangle$$
$$N_1^2 = N_1^1 \sqcup [x \ge 1, N_2^1]^\natural = \{\langle x : (-\infty, 0], y : [0]\rangle, \langle x : [1; 99], y : (-\infty; 0]\rangle\}$$

repeating the steps gives

$$N_1^3 = \{\langle x : (-\infty, 0], y : [0]\rangle, \langle x : [1; 99], y : (-\infty; 0]\rangle\} = N_1^2$$

Thus, we take

$$N_1 = \{\langle x : (-\infty, 0], y : [0]\rangle, \langle x : [1; 99], y : (-\infty; 0]\rangle\}$$
$$N_2 = \langle x : (-\infty; 99], y : (-\infty; 0]\rangle$$
$$N_3 = \langle x : \top, y : (-\infty; 0]\rangle$$
$$N_5 = \top$$
$$N_8 = \{\langle x : \top, y : [0]\rangle, \langle x : [0; 98], y : (-\infty; 0]\rangle\}$$
$$N_9 = \langle x : \top, y : [0]\rangle$$
$$N_{12} = \{\langle x : (-\infty; 0], y : [0]\rangle, \langle x : [0; 98], y : (-\infty; 0]\rangle\}$$
$$N_{14} = \langle x : \top; y : [0]\rangle$$

To initialize the positive side, again we use a universal recurrent set produced for us by E-HSF. The query to E-HSF is shown in Fig. 9. The result is $R_\forall = \langle x : [100; +\infty), y : \top\rangle$.

Again, we choose to use *pre* for all the computation steps on the positive side.

$$P_{14}^1 = pre(\texttt{assert}\ y \neq 0, \top) - N_{14} =$$

$\{\langle x : \top, y : (-\infty; -1]\rangle, \langle x : \top, y : [1; +\infty)\rangle\}$
$P_1^1 = R_\forall \sqcap (D_{x \leq 0} \sqcap P_{15}^1) - N_1^1 =$
$\{\langle x : [100; +\infty), y : \top\rangle, \langle x : (-\infty; 0], y : (-\infty; -1]\rangle, \langle x : (-\infty; 0], y : [1; +\infty)\rangle\}$

Before we proceed to the loop body, we need to make a remark on using the combination of *pre* and subtraction on the positive side. When abstract program is non-deterministic (because of non-determinism of a concrete program or coarseness of abstraction), *pre* is often larger that *wp*, and coarse subtraction can turn it into $\bot$. Consider a fragment of the current example in Fig. 10a. The trivial translation to our input language is shown in Fig. 10b: having $y \leq 0$ allows to execute the assertion (and fail), but it is always possible to skip the assertion. If we try to analyze the fragment of Fig. 10b in isolation, using *pre* for the positive side, we get the following

$$N_5 = \top$$
$$P_5 = \bot$$
$$N_3 = [y \leq 0, N_5]^\natural \sqcup (\top \sqcap \bot) = \langle y : (-\infty; 0]\rangle$$
$$P_3 = [y \leq 0, P_5]^\natural \sqcup (\top \sqcap P_\omega) - N_3 = \top - N_3$$

If we use a course subtraction of (1), we get

$$P_3 = \bot$$

That is, in this case, we lose all of the positive side, even though there are input states for which the program fragment is safe. The problem here is that the precondition for safety ($y \geq 1$) never had a chance to materialize, and (because of the use of *pre*) $\top$ easily got into P_3. To work around this in our simple example, we use the following tricks (and a real-world tool could use some form of trace partitioning [15]). First, we translate conditions with $*$ into nested conditions, as in Fig. 10c or 10d. This allows for $[y \geq 1]$ to appear in the equations. Second, for *some* steps of the computation, we allow for a domain element D to be redundant, i.e., to contain such disjuncts $d_1, d_2 \in D$ that $d_1 \sqsubseteq d_2$. Note that widening operators for power sets may require that the domain elements are not redundant [2], and we would have to remove redundancy before applying widening. Then, from the fragment in Fig. 10d, we get the following (for Fig. 10c, the steps are almost the same):

$$N_5 = \top$$
$$P_5 = \bot$$
$$N_4 = ([y \leq 0, N_5]^\natural \sqcup [y \geq 1, \bot]^\natural) = \langle y : (-\infty; 0]\rangle$$
$$P_4 = ([y \leq 0, P_5]^\natural \sqcup [y \geq 1, \top]^\natural) - N_4 = \langle y : [1; +\infty)\rangle$$
$$N_3 = \bot \sqcup N_4 = \langle y : (-\infty; 0]\rangle$$
$$P_3 = (\top \sqcup P_4) - N_3 = (\top \sqcup \langle y : [1; +\infty)\rangle) - N_3 = \langle y : [1; +\infty)\rangle$$

This way, we were able to show that the fragment is safe for the states in $\langle y : [1; +\infty)\rangle$.

```
3 | if  y ≤ 0 ∧ ∗ then
4 |
5 |   assert 0
6 |
7 | end
```

(a) Informal program.

$_3(([y \le 0];_5 \texttt{assert}\,0) + \texttt{skip})$

(b) Single branching.

$_3(([y \le 0];_4(_5 \texttt{assert}\,0 + \texttt{skip})) + ([y \ge 1]; \texttt{skip}))$

(c) Nested branching (1).

$_3(_4(([y \le 0];_5 \texttt{assert}\,0) + ([y \ge 1]; \texttt{skip})) + \texttt{skip})$

(d) Nested branching (2).

Fig. 10: Representations of non-deterministic branching.

Equipped with these tricks, we return to the example (actually, here we prefer the scheme in Fig. 10c to handle the condition in line 3). Recall that

$$P_1^1 = \{\langle x : [100; +\infty), y : \top\rangle, \langle x : (-\infty; 0], y : (-\infty; -1]\rangle, \langle x : (-\infty; 0], y : [1; +\infty)\rangle\}$$

then

$$P_{12}^1 = \{\langle x : [99; +\infty), y : \top\rangle, \langle x : (-\infty; -1], y : (-\infty; -1]\rangle, \langle x : (-\infty; -1], y : [1; +\infty)\rangle\}$$
$$P_9^1 = \{\langle x : \top, y : (-\infty; -1]\rangle, \langle x : \top, y : [1; +\infty)\rangle\}$$
$$P_8^1 = \langle x : \top, y : [1; +\infty)\rangle$$
$$P_5^1 = \bot$$
$$P_3^1 = \langle x : \top, y : [1; +\infty)\rangle$$
$$P_2^1 = \{\langle x : (-\infty; 99], y : [1; +\infty)\rangle, \langle x : [100; +\infty), y : \top\rangle\}$$
$$P_1^2 = \{\langle x : [100; +\infty), y : \top\rangle, \langle x : (-\infty; 0], y : (-\infty; -1]\rangle, \langle x : (-\infty; 0], y : [1; +\infty)\rangle, \langle x : [1; 99], y : [1; +\infty)\rangle\}$$

then, if we continue in the same way, we get

$$P_1^3 = P_1^2$$

thus, we take

$$P_1 = \{\langle x : [100; +\infty), y : \top\rangle, \langle x : (-\infty; 0], y : (-\infty; -1]\rangle, \langle x : (-\infty; 0], y : [1; +\infty)\rangle, \langle x : [1; 99], y : [1; +\infty)\rangle\}$$

Example 3 (extended) The purpose of the example is to demonstrate how our problem decomposition works for the domain of 3-valued structures.

The tracked instrumentation predicates are listed in Table 1. We use subtraction as defined in (1). First, we produce the approximation of the negative side. In this example,

Table 1: Instrumentation predicates used in shape analysis examples.

Reachability between nodes via n	$t(v, v_1)$	$n^*(v, v_1)$
Reachability from variable x	$r_x(v)$	$\exists v_1.\ x(v_1) \wedge n^*(v_1, v)$
Sharing via n	$is(v)$	$\exists v_1 \neq v_2.\ n(v_1, v) \wedge n(v_2, v)$
Existence of n-successor	$hn(v)$	$\exists v_1.\ n(v, v_1)$

the negative side of the loop entry location (N_1) stabilizes as $\bot$. Informally, this is because the potential source of failure at line 2 is guarded by the loop condition, and the structures that could fail are prevented from entering the loop. We could stop the analysis at this point and report P_1 to be $\top$, but for the purpose of the demonstration, we continue.

Next, we approximate the existential recurrent set for the loop. We implemented a simplistic procedure that unrolls the loop body multiple times and identifies the witnesses to recurrence. A bit more specifically, we build a descending chain of heaps that may enter the loop at least k times:

$$R_1 \sqsupseteq R_2 \sqsupseteq R_3 \sqsupseteq \dots$$

Assuming that ψ is the condition that allows to enter the loop (in this example, $x \neq nil$), φ is the condition that allows to exit the loop (in this example, $x = nil$), the statement right following the loop is labeled by o (in this example, we can assume that it is `skip`), and P_o is already known (in this example, $\top$)

$$R_1 = [\psi, P_o]^\natural \sqcup [\neg\varphi]^\natural$$
$$R_k = ([\psi, P_o]^\natural \sqcup [\neg\varphi]^\natural) \sqcap pre(R_{k-1}),\ \text{for } k \geq 2$$

Then, we rely on the ability of our backward transformers to materialize structures. It would normally be the case that for $k \geq 2$ there is a subset of structures $W \subseteq R_k$ (that we call a witness), s.t.

$$pre^{k-1}(W) \sqsupseteq W$$

We prefer to not use $[\psi]^\natural$ directly as R_1 to rule out the heaps that are known to lead to a failure if they exit the loop (in case ψ and φ are not mutually exclusive).

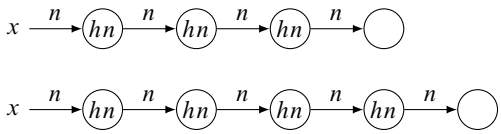
We summarize the witnesses and thus obtain an approximation of the existential recurrent set. In this example, the recurrent set consists of cyclic lists that our procedure is able to summarize in the form of 8 shapes. One of them (probably, giving the best idea of the recurrent set) is shown in Fig. 6. For each node, inside the circle, we list the instrumentation predicates that are evaluate to 1 for it; predicate t_n is not displayed, but its evaluation can be deduced from r_x. The structure in Fig. 6 represents a "panhandle" cyclic list (where the list head is not touched).

The initial approximation of the positive side (for location 1) consists of the structures that immediately exit the loop (without visiting the body), i.e., structures where x does not point to a node. Then, the analysis summarizes all the predecessors of such structures. Our procedure is able to summarize them in 9 shapes, one of which (probably, giving the best idea of the set) is shown in Fig. 7 (it represents an acyclic list of the length 3 or more).

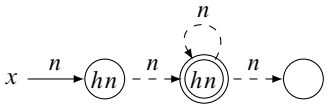
Thus, we were able to identify that both cyclic and acyclic lists are safe inputs for the program.

Example 4 (extended) The example demonstrates what happens if the used abstraction is not precise enough to capture the properties of the positive side. While the first step (at line 2) is still guarded by the loop condition, the second step (at line 3) is a source of failure. That is, the program fails when given a list of odd length as an input.

But our abstraction is not expressive enough to encode such constraints on the length of the list: this is illustrated in Fig. 11 where lists of different length are abstracted to the same bounded structure. As a result, the produced summary of the negative side (for location 1) contains, e.g., the structure in Fig. 7 that represents lists of length just 3 or more, both even and odd.



(a) Concrete lists of even and odd length.



(b) Result of canonical abstraction for both.

Fig. 11: Lists of different lengths abstracted to the same bounded structure.

Informally, we can see that the program terminates successfully when given a list of even length as an input. But applying canonical abstraction would usually remove the information about list length from the structures, and such structures would be removed from the positive side by subtraction. The only finite list that the analysis is able to identify as safe is the list of length exactly two that is shown in Fig. 12.

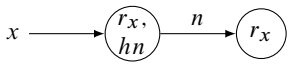


Fig. 12: List of length exactly 2.

The recurrent set for this example again consists of cyclic lists, and the summary that we produce is the same as in Example 3. Our abstraction is precise enough to distinguish cyclic lists from acyclic, and to show that cyclic lists represent safe inputs.

In this example, we can see precision loss resulting from using too coarse abstraction. In this example, the analysis was not able to summarize the precondition for successful termination (as expected), but still was able to produce the summary of the states that cause non-termination without failure.